Document Structure Analysis for the NTCIR-5 Patent Retrieval Task

Atsushi Fujii, Tetsuya Ishikawa
University of Tsukuba

PATENT-12

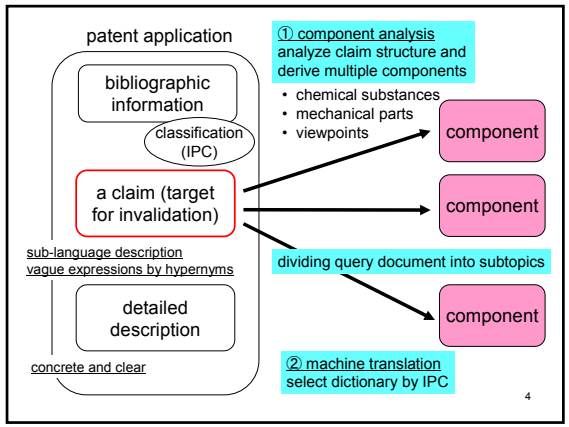
Invalidity search task

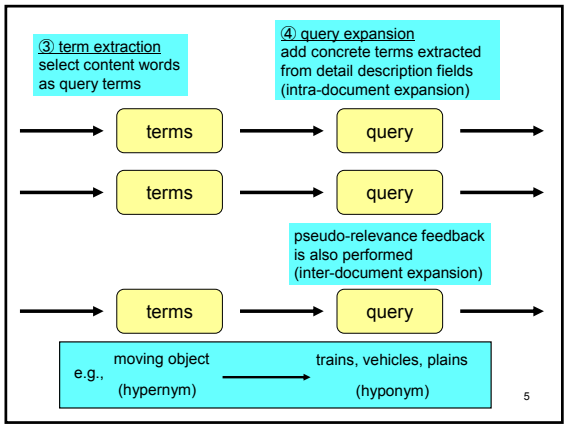
- Find the patents that can invalidate the demand in a patent application (claim)
- This can be seen as patent-to-patent associative retrieval
  - both queries and documents are patents
- This task is usually performed by
  - examiners in a government patent office
  - searchers of IP division in private companies

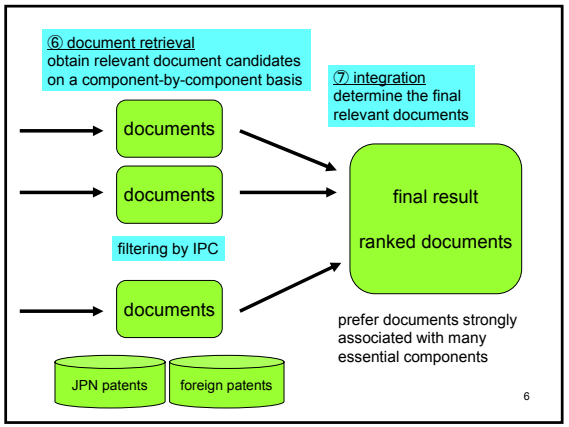
Basis of our system

- Because query is long, structure analysis is effective

1. analyze claim structure and divide query into subtopics (local structure analysis)
2. expand query terms using "detailed description" fields (global structure analysis)
3. search for documents on a subtopic-by-subtopic basis
4. integrate and re-rank documents







Example claim

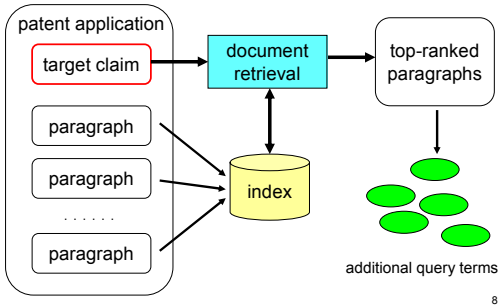
components (subtopics)

	Claim #1	Candidate documents A	B	C
1	移動体の周囲を監視するための移動体の周囲監視装置であって、	100	20	0
2	その周囲の領域の映像を光学像に中心射影変換する光学系と、撮像レンズを含み、前記中心射影変換された光学像を画像データに変換する撮像部とを含む少なくとも1つの全方位視覚センサーと、	40	10	100
3	前記画像データをパノラマ画像データおよび透視画像データの少なくとも一方に変換する画像処理部と、	10	0	30
⋮	⋮	⋮	⋮	⋮
7	前記表示部が、前記移動体の周囲を俯瞰する前記透視画像を表示する、移動体の周囲監視装置。	80	30	0
	weighted average of scores	41.4 (1)	17.1 (3)	27.1 (2)

relevance score

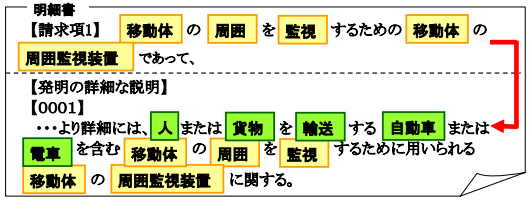
7

Query expansion (intra-doc)



8

Example of intra-doc expansion



9

Comparative experiments

- Optional methods (yes/no)
  - A: component analysis
    - based on Japanese punctuation
  - B: intra-document expansion
  - C: character bigram index terms (combined with word index terms)
  - D: pseudo-relevance feedback
  - E: filtering by IPC

10

Results of Doc IR (MAP)

A	B	C	D	E	N4A	N5A
yes	yes	yes	yes		21.37	19.16
yes	yes	yes	yes	yes	20.84	20.75
yes	yes		yes	yes	19.86	18.50
	yes	yes		yes	21.15	20.18
	yes	yes	yes	yes	19.83	21.07
		yes	yes	yes	20.84	20.75

11

Results of Pas IR (CRS)

A	B	D	CRS
yes	yes	yes	12.01
yes	yes		12.12
	yes	yes	10.91
		yes	11.23
Baseline			16.23

12